Volume 165, number 2 FEBS 1104 January 1984

The complete amino acid sequence of hirudin, a thrombin specific inhibitor

Application of colour carboxymethylation

Johannes Dodt, Hans-Peter Müller+, Ursula Seemüller and Jui-Yoa Chang*

Department of Clinical Chemistry and Clinical Biochemistry, University of Munich, Munich, +Plantorgan Werk KG, Bad Zwischenahn, FRG and **Pharmaceuticals Research Laboratories, CIBA-GEIGY Ltd, Basel, Switzerland*

Received 21 November 1983

Color carboxymethylation of cysteine residues with a new chromophoric reagent dimethylaminoazobenzene iodoacetamide, was applied to the micro-sequence analysis of hirudin, a thrombin specific inhibitor. Six cysteine residues of the reduced hirudin were detected as colored phenylthiohydantoin derivative and 3 tryptic peptides of hirudin (all containing cysteines) were isolated as colored peptide. The complete hirudin sequence, including 6 uncertain positions left in the previous report [Petersen T.E. et al. (1976) in: Protides of the Biological Fluids; 23rd Colloquium, pp. 145, Pergamon Press, London] was established.

Hirudin *Tyrosine sulfate* *Color carboxymethylation*

1. INTRODUCTION

Hirudin is a potent thrombin-specific inhibitor [1,2] isolated from the salivary glands of the leech *Hirudo medicinalis*. It forms a tight complex with α-thrombin with an equilibrium dissociation constant of as low as 0.82×10^{-12} mol/l [3]. This high binding affinity had led to the speculation that interaction of hirudin with α-thrombin must comprise multiple binding sites [4,5]. In an attempt to elucidate the molecular mechanism of thrombin inhibition by hirudin, the complete amino acid sequence has to be determined.

We describe here the complete sequencing of hirudin by applying a newly developed technique [6]. (a) Cysteine was identified as a colored derivative by specific labelling with DABIA. The colored cysteine derivatives are directly visible at the 100 pmol level in the sequenator fractor collector (e.g. residue 39 in fig.1) and can be analysed by HPLC at the 5–10 pmol level. (b) Peptides were isolated as colored peptides by labelling at the cysteine residue. Since every tryptic peptide of hirudin contains at least one cysteine, all tryptic peptides were isolated in colored form without supplementing an additional derivatization method [7,8]. This technical approach has demonstrated that the sequence of a small cysteine-containing protein can be determined in an efficient and sensitive way.

Abbreviations: DABIA, 4-dimethylaminobenzene 4′-iodoacetamide; DABITC, dimethylaminoazobenzene isothiocyanate; DABS-Cl, dimethylaminoazobenzene sulfonyl chloride; HPLC, high-performance liquid chromatography; PTH, phenylthiohydantoin

2. MATERIALS AND METHODS

2.1. *Materials*

Hirudin was isolated from the whole body of leeches as in [1] with some modifications (to be published) and its purity was ascertained by SDS–gel electrophoresis and N-terminal analysis [9]. DABIA was synthesized as in [6]. DABS-Cl

Published by Elsevier Science Publishers B.V.

and DABITC were obtained from Fluka. All other chemicals and solvents used in both HPLC analysis and sequenator run were commercial analytical grade from Merck.

2.2. *Methodologies used in protein chemistry*

Protein was reduced [10] and color carboxymethylated with a 2–3 molar excess of DABIA as in [6]. Excess DABIA was removed by gel filtration (G-25 superfine, eluted with 50% acetic acid). Amino terminal of polypeptide was determined by the DABITC method [9]. Amino acid composition and C-terminal sequence were analysed by the DABS-Cl method [11,12] and carboxypeptidase Y [13]. Automatic sequencing [14] was performed on a modified Beckman 890C sequenator [15,16] and the released phenylthiohydantoins were analyzed by JPLC on a Zorbax-CN column [16]. PTH derivative of DABIA carboxymethylated cysteine was also quantitatively analysed by HPLC [6].

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequencing of hirudin*

Amino acid composition analysis by both standard analyser [17] and the DABS-Cl method [11,12] shows that hirudin contains 3 molecules of lysine and 6 molecules of cysteine per molecule of hirudin. No arginine was detected (not shown).

Native hirudin (60 nmol) was reduced with dithiothreitol [10] and sulfhydryl groups were labelled with DABIA [6]. Two samples (7 and 3.5 nmol) of the colored hirudin were subjected to automatic Edman degradation [14] and in both runs, only the first 39 amino acids were determined. Six cysteine residues were assigned at positions 6, 14, 16, 22, 28 and 39 directly through their visible purple-red color in the fraction collector tubes at the sequenator (fig.1).

About 12 nmol colored hirudin was digested with trypsin and the colored tryptic peptides were isolated by HPLC and detected in the visible region (fig.2). They were recovered (yield 2.5–5 nmol) and sequenced by automatic Edman degradation, all through their C-terminal ends. From the PTH derivative analysis, residue 63 (from peptide in fraction 1 of fig.2) was found to be Tyr instead of Tyr-SO_3H reported in [2]. Fractions 1, 3 and 4 account for the complete sequence of hirudin and the positions of the 6 colored cysteines were

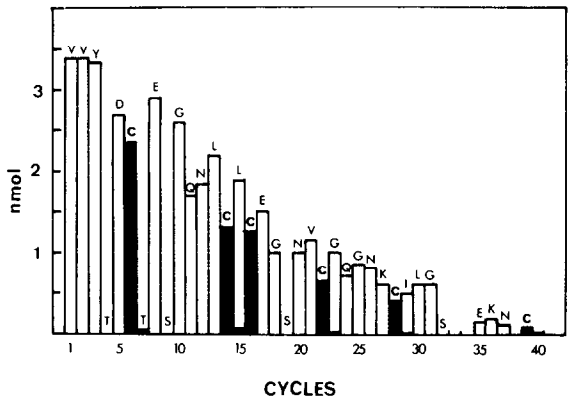

Fig.1. Yields of PTH derivatives of amino acids from a sequenator run performed with 3.5 nmol hirudin. Both Ser and Thr were positively identified through their characteristic by-products [15] but were not quantitatively analysed. Phenylthiohydantoins of color-labelled cysteine were quantitatively analysed at 436 nm filled columns). The first 39 amino acids were identified but uncertain gaps were left at positions 33, 34 and 38. Those gaps were finally assigned by sequencing the tryptic peptides.

once more confirmed in these tryptic peptides.

The final stage of hirudin sequencing was C-terminal sequence analysis. One nmol carboxymethylated hirudin was digested by carboxypeptidase Y and released amino acids were analysed by the DABS-Cl method [11,12] (fig.3). The digestion drastically slowed down at residue 58 and although nearly equal amounts (mol) of Gln, leu, Pro, Ile and double (mol) of Glu were released after 20 min digestion, only 0.12 mol of Tyr was recovered. Instead, there appeared an unknown acidic amino acid (fig.3). This result did not fit into the sequence obtained from tryptic peptide 1, as automatic sequencing clearly indicated that position 63 was Tyr. It was soon realized that this unknown acidic amino acid was a Tyr derivative which can be readily converted to Tyr by mild acid hydrolysis. This Tyr derivative was identified as Tyr-O-sulfate through the following observations and experiments: (a) Its marked acidic lability is characteristic for an aryl sulfate [18]. (b) This acidic Tyr derivative is not Tyr-O-phosphate, since a standard Tyr-O-phosphate (Sigma) is eluted immediately after Glu. (c) This acidic Tyr derivative has the same elution time as synthetic Tyr-O-sulphate (synthesized as in [19]). The C-terminal sequence

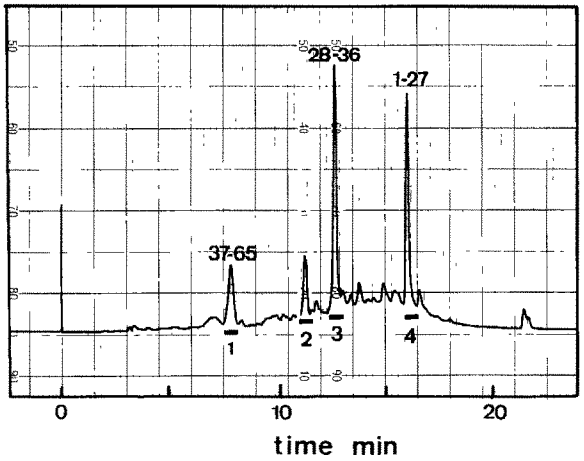

Fig.2. Preparative isolation of hirudin-derived cysteine-containing peptides by HPLC. The DABIA labelled hirudin (12 nmol) was digested by trypsin in 150 μl of 0.1 M ammonium bicarbonate at 37°C for 24 h. Fifty percent (75 μl) of the sample (6 nmol) was loaded for each preparative isolation run. Recoveries were 20–40%. The peptides in fractions 1, 3 and 4 account for the complete sequence of hirudin and their corresponding positions in the hirudin sequence are indicated. No sequence result was obtained from automatic sequencing of peptide in fraction 2. Chromatographical conditions were: solvent A, 0.01 M phosphate (pH 7.0), containing 2% dimethylformamide; solvent B, acetonitrile; gradient, 20–75% B in 18 min; column, Merck RP-18 (5 μm); column temperature, 22°C; Detector, 436 nm, 2.0 absorption unit full scale.

Table 1

Time-course release of C-terminal amino acids from hirudin by carboxypeptidase Y digestion[a]

Period of digestion (min)	- Ile	- Pro	- Glu[b]	- Glu[b]	- Tyr[c]	- Leu	- Gln
3	0.12	0.19	(0.24)	0.50	0.61	0.95	
10	0.42	0.54	(0.66)	0.87	0.91	0.97	
20	0.73	0.89	(1.08)	0.99	0.98	0.96	

[a] Aliquots of 10 pmol were used for amino acid analysis. The listed values represent mol amino acid/mol protein

[b] The listed values represent only 50% of the total Glu yield

[c] The yield of Tyr-O-sulfate was corrected for its 12% desulfated derivative

determined by digestion with carboxypeptidase Y fits precisely with the sequence obtained from Edman degradation (table 1) and the complete hirudin sequence is shown in fig.4.

3.2. *Color carboxymethylation*

Cysteine residues of polypeptides have to be chemically modified prior to sequence determination. There are a number of chemical methods available for this purpose [20,21], including the

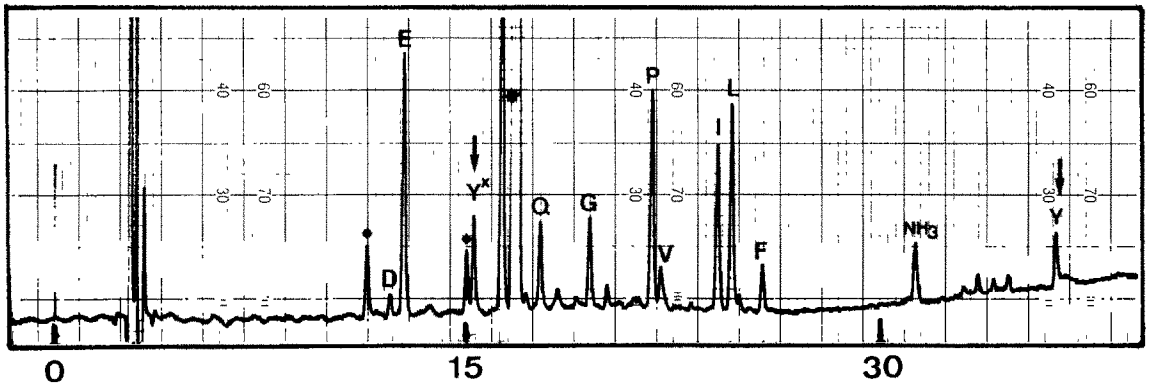

Fig.3. Analysis of amino acids (by DABS-Cl method) released from hirudin after 20 min of carboxypeptidase Y digestion. Two hundred pmol carboxymethylated hirudin were digested with 0.2 μg carboxypeptidase Y in 10 μl phosphate buffer (pH 5.4) at room temperature for 20 min. The sample was freeze-dried, derivatized with DABS-Cl and 5% of the total sample (10 pmol) was applied for amino acid analysis. DABS-amino acids are symbolized by the one letter code of their corresponding amino acid. Y*, DABS-Tyr-O-sulfate. By-products originating from the reagent are marked by asterisk. The appearance of Gly was not expected from the C-terminal sequence of hirudin and is believed to be introduced into the sample during the carboxymethylation. Chromatographical conditions were: solvent A, 0.012 M phosphate (pH 6.5), solvent B, acetonitrile containing 4% dimethylformamide; gradient, 12–37% B/0–22 min, 37–60% B/22–35 min, kept at 60% B at 35–40 min then 60–12% B/40–45 min; column, Merck RP-18 (5 μm); column temperature, 50°C; detector, 436 nm, 0.01 absorption until full scale.

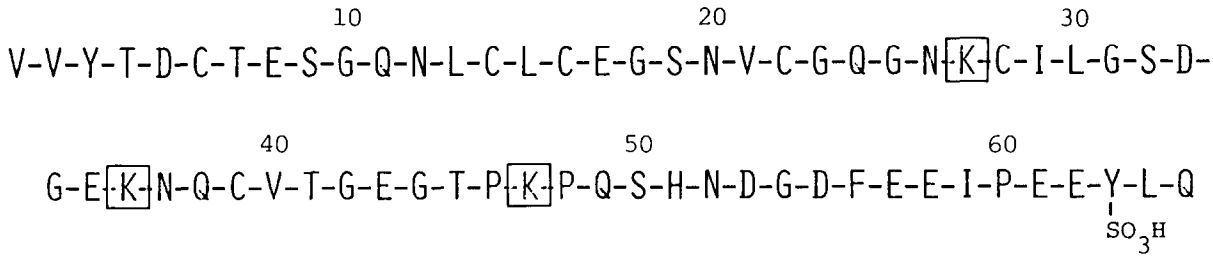

Fig.4. The complete amino acid sequence of native hirudin. By trypsin digestion, the lysyl bonds in positions 27 and 36 were cleaved whereas the lysyl bond in position 47 was resistant.

most popular techniques, such as carboxymethylation [22] and performic acid oxidation [23]. However, analysis of chemically modified cysteine is, in general, more difficult than that of common amino acids. For instance, PTH-cysteic acid is not extractable by organic solvent in the liquid phase Edman degradation and PTH-carboxymethyl cysteine has the tendency of breaking down like the serine derivative. To circumvent this problem in microsequencing, radioactive carboxymethylation became one of the most commonly used approaches.

The color carboxymethylation employing DABIA has advantages over the radioactive carboxymethylation. (a) DABIA labelled cysteine and peptide derivatives [6] have a molar extinction coefficient of 30000. These colored amino acid and peptide derivatives are separated by HPLC and detected in the visible region with the signal: baseline noise ratio of at least 3000 $nmol^{-1}$, whereas radioactive carboxymethylation employing standard conditions [24] gave about 150–200 $nmol^{-1}$ of signal: background noise ratio (assuming the background radioactivity is 20 cpm). (b) DABIA-labelled cysteine is much more hydrophobic than iodoacetic acid-labelled cysteine or cysteic acid, assuring quantitative extraction of cysteine derivative from the spinning cup during microsequence analysis.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Professor H. Fritz and the technical assistance of Mr R. Knecht.

REFERENCES

[1] Bagdy, D., Barabas, E. and Graf, L. (1973) Thromb. Res. 2, 229–238.
[2] Bagdy, D., Barabas, E., Graf, L., Ellebaek, T. and Magnusson, S. (1976) Methods Enzymol. 45, 669–678.
[3] Markwardt, F. and Walsmann, P. (1958) Hoppe-Seyler's Z. Physiol. Chem. 312, 85–98.
[4] Magnusson, S. (1971) in: The Enzymes (Bayer, P.D. ed.) 3rd edn. vol. 3, pp. 277–321.
[5] Fenton, J.W. (1981) Ann. N.Y. Acad. Sci. 267, 469–495.
[6] Chang, J.Y., Knecht, R. and Braun, D.G. (1983) Biochem. J. 211, 163–171.
[7] Chang, J.Y. (1981) Biochem. J. 199, 537–546.
[8] Chang, J.Y., Knecht, R., Ball, R., Alkan, S.S. and Braun, D.G. (1982) Eur. J. Biochem. 127, 625–629.
[9] Chang, J.Y. (1983) Methods Enzymol. 91, 79–84.
[10] Cleland, W.W. (1964) Biochemistry 3, 480–485.
[11] Chang, J.Y., Knecht, R. and Braun, D.G. (1983) Methods Enzymol. 91, 41–48.
[12] Chang, J.Y., Knecht, R. and Braun, D.G. (1982) Biochem. J. 203, 803–806.
[13] Hayashi, R. (1977) Methods Enzymol. 47, 84–93.
[14] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[15] Wittman-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[16] Knecht, R., Seemüller, U., Liersch, M., Fritz, H., Braun, D.G. and Chang, J.Y. (1983) Anal. Biochem. 130, 65–71.
[17] Spackman, D.H., Moore, S. and Stein, W.H. (1958) Anal. Chem. 30, 1190–1205.
[18] Bettelheim, F.R. (1954) J. Am. Chem. Soc. 76, 2838–2839.
[19] Reitz, H.C., Ferrel, R.E., Fraenkel-Conrat, H. and Olcott, H.S. (1946) J. Am. Chem. Soc. 68, 1024–1030.

[20] Scaffone, E. and Fontana, A. (1975) in: Protein Sequence Determination (Neadleman, S.B. ed.) pp. 162–203.
[21] Kenyon, G.L. and Bruice, T.W. (1977) Methods Enzymol. 47, 407–430.
[22] Gurd, F.R.N. (1972) Methods Enzymol. 25, 424–437.
[23] Neumann, N.P. (1972) Methods Enzymol. 25, 393–400.
[24] O'Donnell, I.J., Frangione, B. and Potter, R.R. (1970) Biochem. J. 116, 261–268.